TELOGY
NETWORKS

Voice Over Packet

White Paper

Edward B. Morgan

Telogy Networks, Inc.

20250 Century Boulevard

Germantown, Maryland 20874

301 515 8580 voice I 301 515 7954 fax

http://www.telogy.com

Introduction

Organizations around the world want to reduce rising communications costs. The consolidation of separate voice and data networks offers an opportunity for significant savings. Accordingly, the challenge of integrating voice and data networks is becoming a rising priority for many network managers. Organizations are pursuing solutions which will enable them to take advantage of excess capacity on broadband networks for voice and data transmission, as well as utilize the Internet and company Intranets as alternatives to costlier mediums.

A Voice Over Packet application meets the challenges of combining legacy voice networks and packet networks by allowing both voice and signaling information to be transported over the packet network. This paper references a general class of packet networks since the modular software objects allow networks such as ATM, Frame Relay, and Internet/Intranet (IP) to transport voice. An overview of a software architecture utilizing Embedded Communication Objects™ (ECOs™) that support Voice Over Packet applications is presented.

ECOs are real-time software and hardware modules that can be dynamically configured to provide flexibility and scalability in communication systems. ECOs have well defined Application Programming Interfaces (APIs), support multiple channel operation through the use of an instance structure, and use a dynamic binding mechanism that allows flexibility in system configuration. Customers can gain a considerable advantage in time to market by using ECOs in building their communication systems.

As shown in Figure 1, the legacy telephony terminals that are addressed range from standard two wire Plain Old Telephone Service (POTS) and Fax Terminals to digital and analog PBX interfaces. Packet networks supported are ATM, Frame Relay, and Internet.
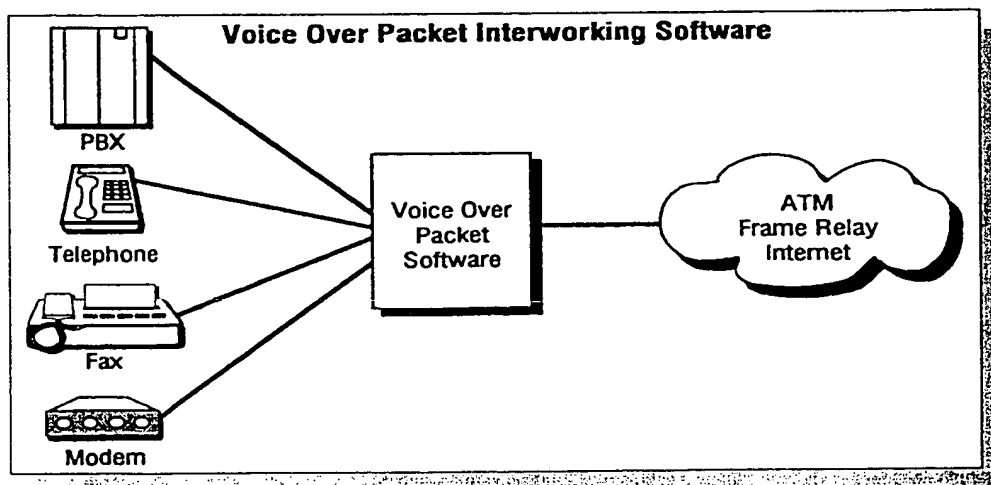

Figure 1

Applications

A wide variety of applications are enabled by the transmission of Voice Over Packet networks. This paper will explore three examples of these applications.

The first application, shown in Figure 2, is a network configuration of an organization with many branch offices (e.g. a bank) that wants to reduce costs and combine traffic to provide voice and data access to the main office. This is accomplished by using a packet network to provide standard data transmission while at the same time enhancing it to carry voice traffic along with the data. Typically, this network configuration will benefit if the voice traffic is compressed due to the low bandwidth available for this access application. Voice Over Packet provides the Interworking Function (IWF), which is the physical implementation of the hardware and software that allows the transmission of combined voice and data over the packet network. The interfaces the IWF must support in this case are analog interfaces which directly connect to telephones or Key systems.

The IWF must emulate the functions of both a PBX for the telephony terminals at the branches, as well as the functions of the telephony terminals for the PBX at the home office. The IWF accomplishes this by implementing signaling software that performs these functions.
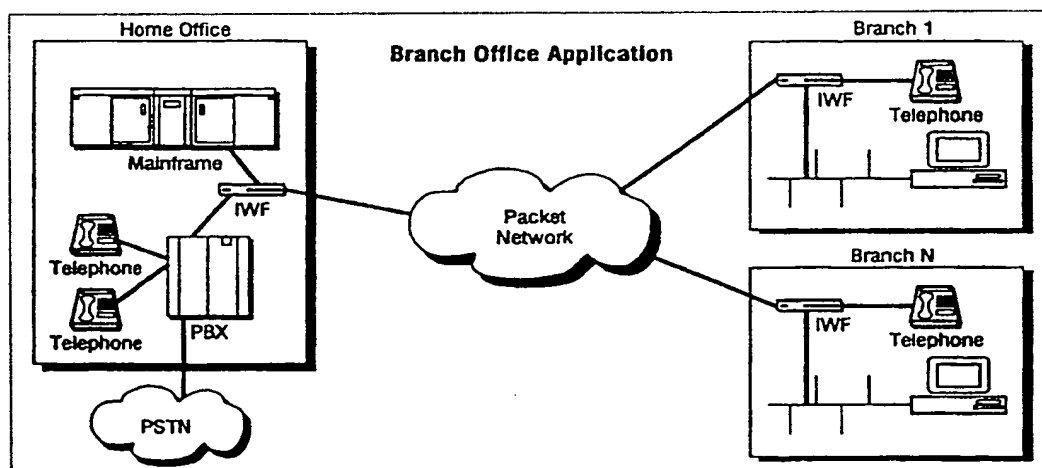

Figure 2

A second application of Voice Over Packet, shown in Figure 3, is a trunking application. In this scenario, an organization wants to send voice traffic between two locations over the packet network and replace the Tie Trunks used to connect the PBXs at the locations. This application usually requires the Interworking Function to support a higher capacity digital channel than the branch application, such as a T1/E1 interface of 1.544 or 2.048 Mbps. The Interworking Function emulates the signaling functions of a PBX, resulting in significant savings to companies' communications costs.
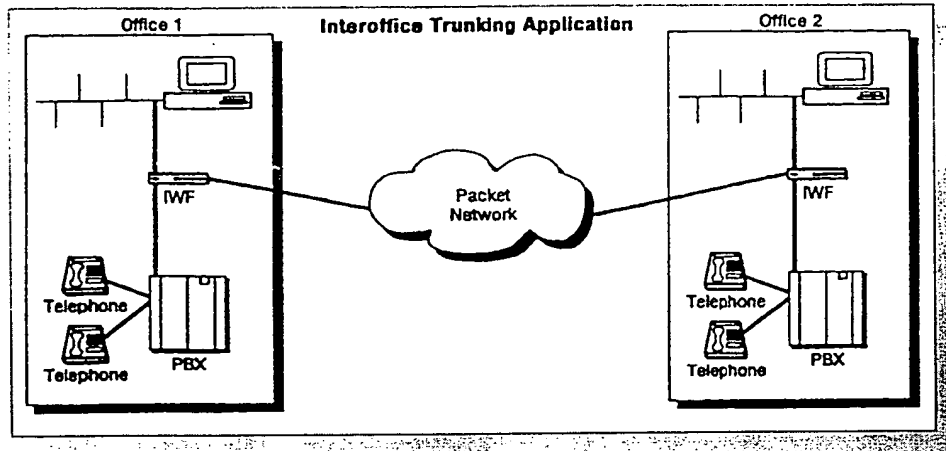
Figure 3

A third application of Voice Over Packet software is interworking with Cellular Networks, as shown in Figure 4. The voice data in a digital cellular network is already compressed and packetized for transmission over the air by the cellular phone. Packet networks can then transmit the compressed cellular voice packet, saving a tremendous amount of bandwidth. The IWF provides the transcoding function required to convert the cellular voice data to the format required by the Public Switched Telephone Network (PSTN).
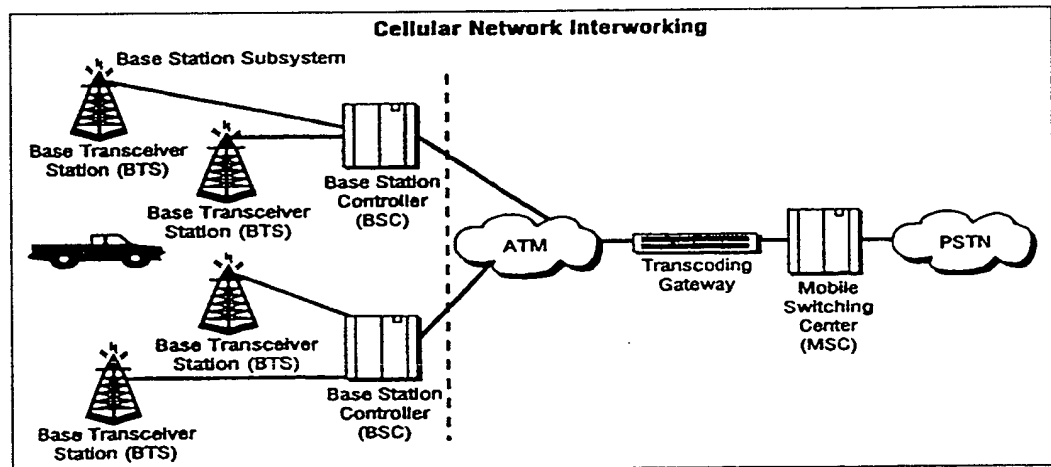

Figure 4

Quality of Service

The advantages of reduced cost and bandwidth savings of carrying Voice Over Packet networks are associated with some quality of service issues unique to packet networks. These issues are explored below.

Delay

Delay causes two problems — echo and talker overlap.

Echo is caused by the signal reflections of the speaker's voice from the far end telephone equipment back into the speaker's ear. Echo becomes a significant problem when the round trip delay becomes greater than 50 milliseconds. Since echo is perceived as a significant quality problem, Voice Over Packet systems must address the need for echo control and implement some means of echo cancellation.

Talker overlap (or the problem of one talker stepping on the other talker's speech) becomes significant if the one-way delay becomes greater than 250 msec. The end-to-end delay budget is therefore the major constraint and driving requirement for reducing delay through a packet network.

Following are sources of delay in an end to end Voice Over Packet call:

1. **Accumulation Delay** (sometimes called algorithmic delay): This delay is caused by the need to collect a frame of voice samples to be processed by the voice coder. It is related to the type of voice coder used and varies from a single sample time (.125 microseconds) to many milliseconds. A representative list of standard voice coders and their frame times follows:

    - ▧ G.726 ADPCM (16, 24, 32, 40 Kbps) - .125 microseconds
    - ▧ G.728 - LD-CELP(16 Kbps) - 2.5 milliseconds
    - ▧ G.729 - CS- ACELP (8 Kbps) - 10 milliseconds
    - ▧ G.723.1 - Multi Rate Coder (5.3, 6.3 Kbps) - 30 milliseconds

2. **Processing Delay:** This delay is caused by the actual process of encoding and collecting the encoded samples into a packet for transmission over the packet network. The encoding delay is a function of both the processor execution time and the type of algorithm used. Often, multiple voice coder frames will be collected in a single packet to reduce the packet network overhead. For example, three frames of G.729 codewords, equaling 30 milliseconds of speech, may be collected and packed into a single packet.

3. **Network Delay:** This delay is caused by the physical medium and protocols used to transmit the voice data, and by the buffers used to remove packet jitter on the receive side. Network delay is a function of the capacity of the links in the network and the processing that occurs as the packets transit the network. The jitter buffers add delay which is used to remove the packet delay variation that each packet is subjected to as it transits the packet network. This delay can be a significant part of the overall delay since packet delay variations can be as high as 70-100 msec in some Frame Relay networks and IP networks.

Jitter

The delay problem is compounded by the need to remove jitter, a variable inter-packet timing caused by the network a packet traverses. Removing jitter requires collecting packets and holding them long enough to allow the slowest packets to arrive in time to be played in the correct sequence. This causes additional delay.

The two conflicting goals of minimizing delay and removing jitter have engendered various schemes to adapt the jitter buffer size to match the time varying requirements of network jitter removal. This adaptation has the explicit goal of minimizing the size and delay of the jitter buffer, while at the same time preventing buffer underflow caused by jitter.

Two approaches to adapting the jitter buffer size are detailed below. The approach selected will depend on the type of network the packets are traversing.

1. The first approach is to measure the variation of packet level in the jitter buffer over a period of time, and incrementally adapt the buffer size to match the calculated jitter. This approach works best with networks that provide a consistent jitter performance over time, such as ATM networks.

2. The second approach is to count the number of packets that arrive late and create a ratio of these packets to the number of packets that are successfully processed. This ratio is then used to adjust the jitter buffer to target a predetermined allowable late packet ratio. This approach works best with the networks with highly variable packet inter-arrival intervals, such as IP networks.

In addition to the techniques described above, the network must be configured and managed to provide minimal delay and jitter, enabling a consistent quality of service.

Lost Packet Compensation

Lost packets can be an even more severe problem, depending on the type of packet network that is being used. Because IP networks do not guarantee service, they will usually exhibit a much higher incidence of lost voice packets than ATM networks. In current IP networks, all voice frames are treated like data. Under peak loads and congestion, voice frames will be dropped equally with data frames. The data frames, however, are not time sensitive and dropped packets can be appropriately corrected through the process of retransmission. Lost voice packets, however, cannot be dealt with in this manner.

Some schemes used by Voice Over Packet software to address the problem of lost frames are:

1. Interpolate for lost speech packets by replaying the last packet received during the interval when the lost packet was supposed to be played out. This scheme is a simple method that fills the time between non-contiguous speech frames. It works well when the incidence of lost frames is infrequent. It does not work very well if there are a number of lost packets in a row or a burst of lost packets.

2. Send redundant information at the expense of bandwidth utilization. The basic approach replicates and sends the nth packet of voice information along with the (n+1)th packet. This method has the advantage of being able to exactly correct for the lost packet. However, this approach uses more bandwidth and also creates greater delay.

3. A hybrid approach uses a much lower bandwidth voice coder to provide redundant information carried along in the (n+1)th packet. This reduces the problem of the extra bandwidth required, but fails to solve the problem of delay.

Echo Compensation

Echo in a telephone network is caused by signal reflections generated by the hybrid circuit that converts between a 4-wire circuit (a separate transmit and receive pair) and a 2-wire circuit (a single transmit and receive pair). These reflections of the speaker's voice are heard in the speaker's ear. Echo is present even in a conventional circuit switched telephone network. However, it is acceptable because the round trip delays through the network are smaller than 50 msec. and the echo is masked by the normal side tone every telephone generates.

Echo becomes a problem in Voice Over Packet networks because the round trip delay through the network is almost always greater than 50 msec. Thus, echo cancellation techniques are always used. ITU standard G.165 defines performance requirements that are currently required for echo cancellers. The ITU is defining much more stringent performance requirements in the G.IEC specification.

Echo is generated toward the packet network from the telephone network. The echo canceller compares the voice data received from the packet network with voice data being transmitted to the packet network. The echo from the telephone network hybrid is removed by a digital filter on the transmit path into the packet network.

Software Architecture

$\mathbf{T}$wo major types of information must be handled in order to interface telephony equipment to a packet network — voice and signaling information.

As shown in Figure 5, Voice Over Packet software interfaces to both streams of information from the telephony network and converts them to a single stream of packets transmitted to the packet network.
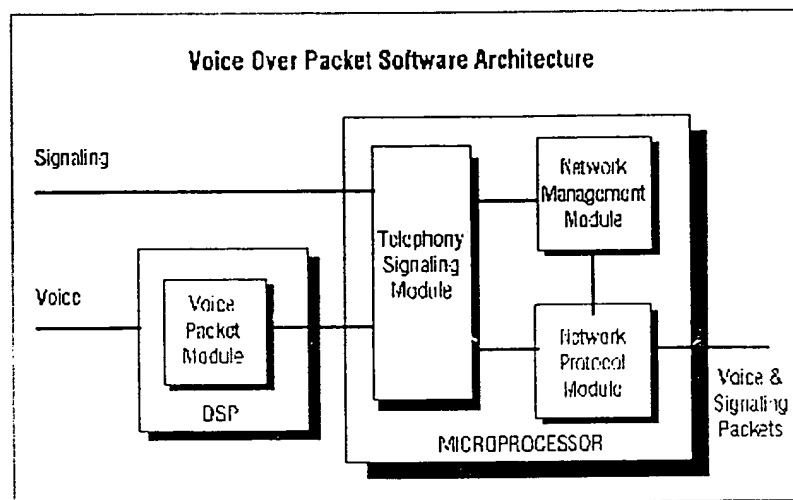

Figure 5

The software functions are divided into four general areas:

1. **Voice Packet Module:** This software, typically run on a DSP, prepares voice samples for transmission over the packet network. Its components perform tone detection and generation, echo cancellation, voice compression, voice activity detection, jitter removal, resampling, and voice packetization.

2. **Telephony Signaling Gateway Module:** This software interacts with the telephony equipment, translating signaling into state changes used by the Network Protocol Module (described below) to set up connections. These state changes are on-hook, off-hook, trunk seizure, etc. This software supports E&M (wink, delay and immediate), Loop or Ground Start FXS and FXO, ISDN BRI/PRI and QSIG.

3. **Network Protocol Module:** This module processes signaling information and converts it from the telephony signaling protocols to the specific packet signaling protocol used to set up connections over the packet network (e.g. Q.933 and Voice over FR signaling). It also adds protocol headers to both voice and signaling packets before transmission into the packet network.

4. **Network Management Module:** This module provides the voice management interface to configure and maintain the other modules of the Voice

7

Over Packet system. All management information is defined in ASN.1 and complies with SNMP V1 syntax. A proprietary Voice Packet MIB is supported until standards evolve in the Forums.

The software is partitioned to provide a well defined interface to the DSP software usable for multiple voice packet protocols and applications. The DSP processes voice data and passes voice packets to the microprocessor with generic voice headers. The microprocessor is responsible for moving voice packets and adapting the generic voice headers to the specific Voice Packet Protocol that is called for by the application, such as Real Time Protocol (RTP), Voice over Frame Relay (VOFR), and Voice Telephony over ATM (VTOA). The microprocessor also processes signaling information and converts it from supported telephony signaling protocols to the packet network signaling protocol (e.g. H.323 (IP), Frame Relay, or ATM signaling).

This partitioning provides a clean interface between the generic voice processing functions — such as compression, echo cancellation, and voice activity detection — and the application specific signaling and voice protocol processing.

Voice Packet Module

This section describes the functions performed by the software in the Voice Packet Module, which is primarily responsible for processing the voice data. This function is usually performed in a Digital Signal Processor (DSP).

The Voice Packet Module consists of the following software:

■ **PCM Interface:** Receives PCM samples from the digital interface and forwards them to appropriate DSP software modules for processing. Forwards processed PCM samples received from various DSP software modules to the digital interface. Performs continuous phase resampling of output samples to the digital interface to avoid sample slips.

■ **Tone Generator:** Generates DTMF tones and call progress tones under command of the Host (e.g. telephone, fax, modem, PBX or telephone switch). Configurable for support of U.S. and international tones.

■ **Echo Canceller:** Performs G.165 compliant echo cancellation on sampled, full-duplex voice port signals. Programmable range of tail lengths.

■ **Voice Activation Detector/Idle Noise Measurement:** Monitors the received signal for voice activity. When no activity is detected for the configured period of time, the software informs the Packet Voice Protocol. This prevents the encoder output from being transported across the network when there is silence, resulting in additional bandwidth savings. This software also measures the Idle Noise characteristics of the telephony interface. It reports this information to the Packet Voice Protocol in order to relay this information to the remote end for noise generation when no voice is present.

8

- **Tone Detector:** Detects the reception of DTMF tones and performs voice/fax discrimination. Detected tones are reported to the Host so that the appropriate speech or fax functions are activated.

- **Voice Codec Software:** Compresses the voice data for transmission over the packet data. Capable of numerous compression ratios through the modular architecture. A compression ratio of 8:1 is achievable with the G.729 voice codec (thus, the normal 64 Kbps PCM signal is transmitted using only 8 Kbps).

- **Fax Software:** Performs a Fax Relay function by demodulating PCM data, extracting the relevant information, and packing the fax line scan data into frames for transmission over the packet network. Significant bandwidth savings can be achieved by this process.

- **Adaptive Playout Unit:** Buffers voice packets received from the packet network and sends them to the Voice Codec for playout. The following features are supported:

  - A FIFO buffer that stores voice codewords before playout removes timing jitter from the incoming packet sequence.

  - A continuous-phase resampler that removes timing frequency offset without causing packet slips or loss of data for voice or voiceband modem signals.

  - A timing jitter measurement which allows adaptive control of FIFO delay.

  The voice packetization protocols use a sequence number field in the transmit packet stream to maintain temporal integrity of voice during playout. Using this approach, the transmitter inserts the contents of a free-running, modulo-16 packet counter into each transmitted packet, allowing the receiver to detect lost packets and to properly reproduce silence intervals during playout.

- **Packet Voice Protocol:** Encapsulates compressed voice and fax data for end-to-end transmission over a backbone network between two ports.

- **Message Processing Unit:** Coordinates the exchange of Monitor and Control information between the DSP and Host via a mailbox mechanism. Information exchanged includes software downline load, configuration data, and status reporting.

- **Real-Time Portability Environment:** Provides the operating environment for the software residing on the DSP. Provides synchronization functions, task management, memory management, and timer management.

Figure 6 diagrams the architecture of the DSP software. The DSP software processes PCM samples from the telephony interface and converts them to a digital format suitable for transmission through a packet network.
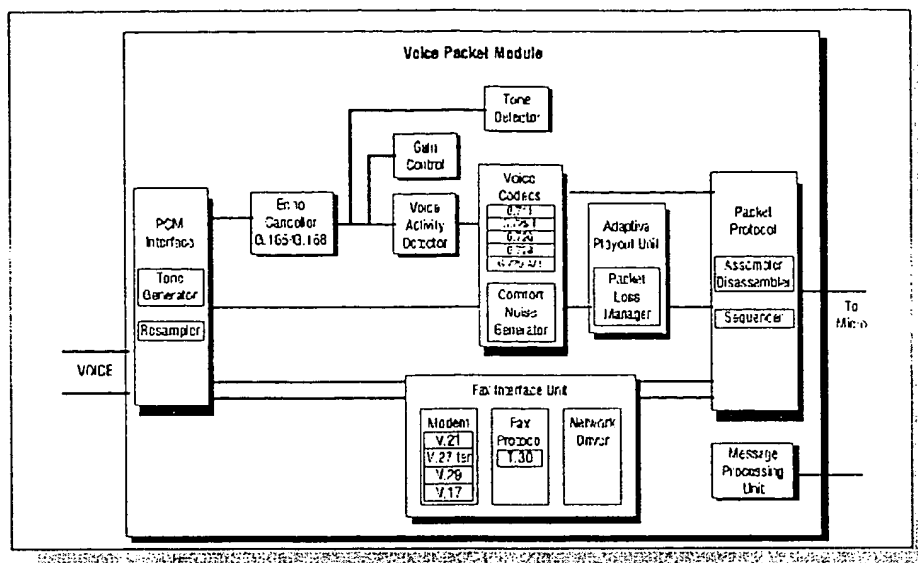
Figure 6

Signaling, Protocol, and Management Modules

The Voice Over Packet software performs telephony signaling to detect the presence of a new call and to collect address (dial digit) information which is used by the system to route a call to a destination port. It supports a wide variety of telephony signaling protocols and can be adaptable to many environments. The software and configuration data for the voice card can be downloaded from a network management system to allow customization, easy installation, and remote upgrades.

The software interacts with the DSP for tone detection and generation as well as mode of operation control based on the line supervision, and interacts with the telephony interface for signaling functions. The software receives configuration data from the network management agent and utilizes operating system services.

Figure 7 diagrams the architecture of the signaling software. The software consists of the following components:
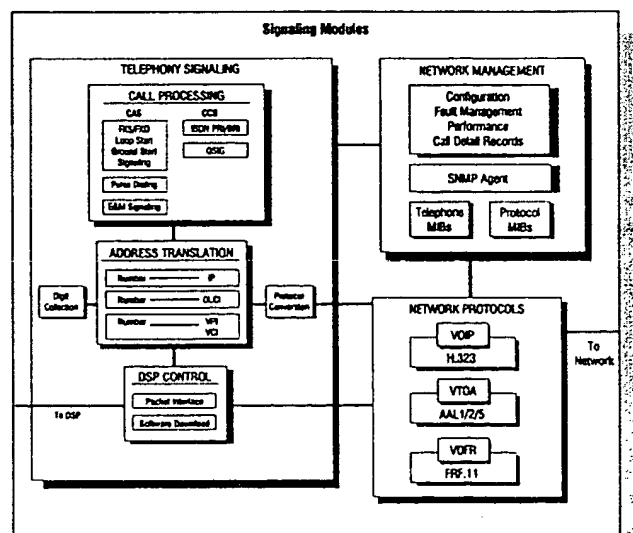

Figure 7

Telephony Signaling Gateway Module

- **Telephony Interface Unit Software:** Periodically monitors the signaling interfaces of the module and provides basic debouncing and rotary digit collection for the interface.

- **Signaling Protocol Unit:** Contains the state machines implementing the various telephony signaling protocols such as E&M.

- **Network Control Unit:** Maps telephony signaling information into a format compatible with the packet voice session establishment signaling protocol.

- **Address Translation Unit:** Maps the E.164 dial address to an address that can be used by the packet network (e.g. an IP address or a DLCI for a Frame Relay Network).

- **DSP Interface Driver:** Relays control information between the Host microprocessor and DSPs.

- **DSP Downline Loader:** Responsible for downline load of the DSPs at start-up, configuration update, or mode changes (e.g. switching from voice mode to fax mode when fax tones are detected).

Network Protocol Module

- **IP Signaling Stack:** H.323 call control and transport software including H.225, H.245 and RTP/RTCP transport protocol, TCP, IP, UDP protocols.

- **ATM Signaling Protocol Stack:** ATM Forum VTOA Voice Encapsulation Protocol. ATM Forum compliant User-Network Interface (UNI) signaling protocol stack for establishing, maintaining, and clearing point-to-point and point-to-multipoint switched virtual connections (SVCs).

- **Frame Relay Protocol Stack:** Frame Relay Forum VOFR Voice Encapsulation Protocol, PVC and SVC Support, Local Management Interface (LMI), Congestion Management and Traffic Monitoring, CIR Enforcement and Congestion.

Network Management Module

The Network Management software consists of three major services addressed in the MIB:

- Physical interface to the telephone endpoint.

- Voice channel service for:
  - processing signaling on a voice channel
  - converting between PCM samples and compressed voice packets

- Call control service for parsing call control information and establishing calls between telephony endpoints.

The Voice Over Packet software is configured and maintained through the use of a proprietary Voice Service MIB.

Summary

A Voice Over Packet software architecture using Embedded Communication Objects (ECOs) has been described for the interworking of legacy telephony systems and packet networks. Some of the key features enabling this application to function successfully are:

- an approach that minimizes the effects of delay on voice quality.

- an adaptive playout to minimize the effect of jitter.

- features that address lost packet compensation and echo cancellation.

- a flexible DSP system architecture that manages multiple channels per single DSP.

Carrying Voice Over Packet networks provides the most bandwidth efficient method of integrating these divergent technologies. While the challenges to this integration are substantial, the potential savings make the investment in a quality implementation compelling.

ABOUT TELOGY NETWORKS

Telogy Networks is the leading provider of embedded communications software to global equipment manufacturers. Telogy's Golden Gateway™ Voice over IP software enables manufacturers to develop connected products that can send real-time voice, fax, and data over multiple packet networks (such as Internet/Intranet, Frame Relay, and ATM). As one of the few embedded software companies with both DSP and microprocessor expertise, Telogy Networks offers its customers truly comprehensive product solutions.

TELOGY *The Leading Provider of Embedded Communications Software*
NETWORKS

20250 Century Boulevard | Germantown, MD 20874 USA | 301 515 8580 voice
301 515 7954 fax | TN@telogy.com | http://www.telogy.com